RESEARCH DATA MANAGEMENT:
WHAT'S INVOLVED?

Jill McTaggart & Sarah Brown

Research Outputs & Impact Team

Scholarly Communication & Repository Services

data@library.uq.edu.au

The University Of Queensland AUSTRALIA | LIBRARY *Enriching world-class scholarship*

TODAY'S OBJECTIVES

- The importance of research data
  - Accountability
  - Efficiency
- How it should be managed
  - What is data
  - Data life cycle
  - Data management issues
  - UQ Library Data Management services
- Where you can get help

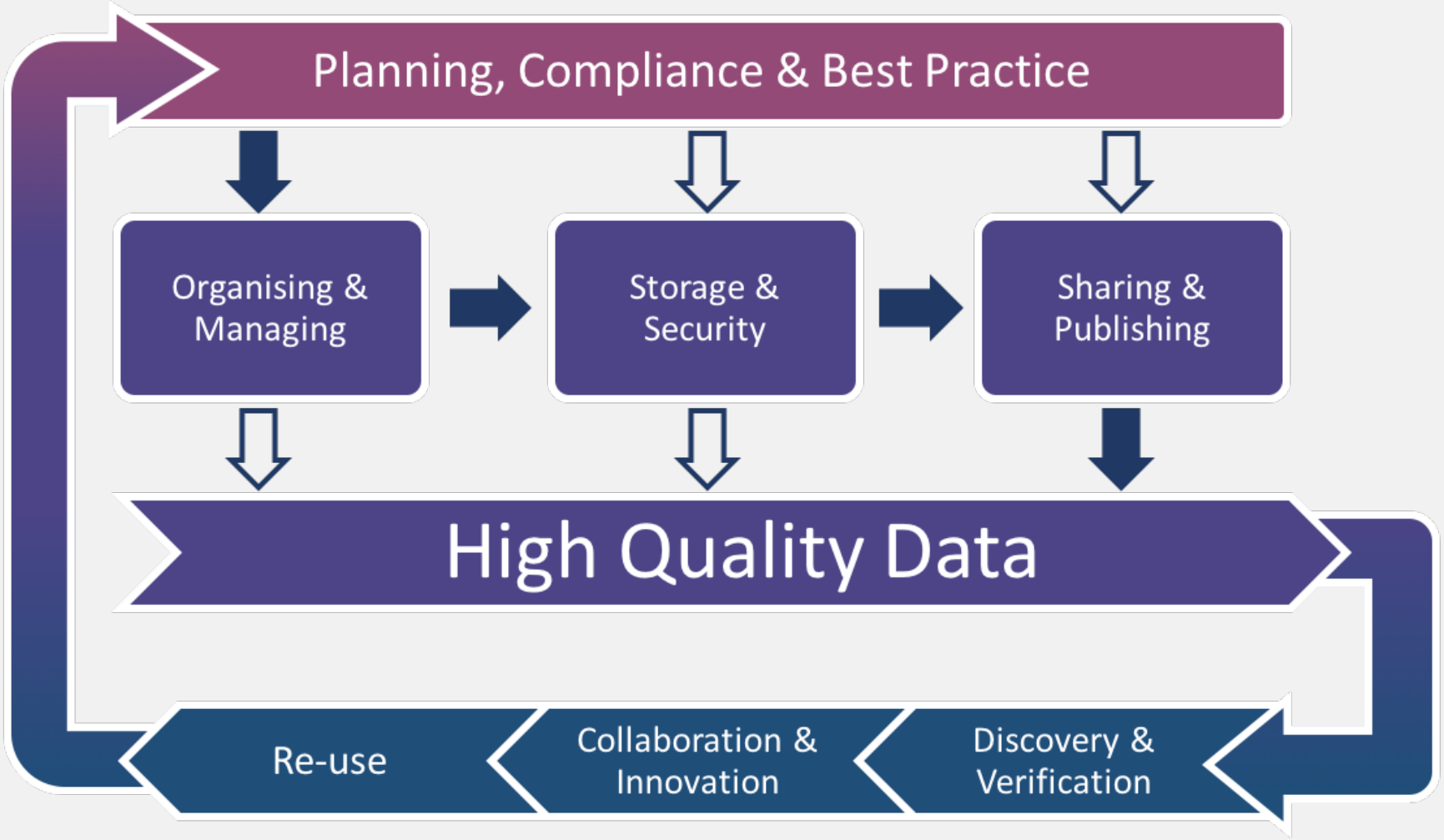
OVERVIEW

RESEARCH INTEGRITY AND RDM

- Accountability and efficiency
- Validate your research
- Avoid allegations of misconduct
- Other benefits of RDM
  - Re-use of research data increases efficiency, reduces duplication of effort
  - Increase in citations for researchers who share data

WHAT IS RESEARCH DATA MANAGEMENT?

- Is a cornerstone of responsible conduct of research
- Involves maintaining, preserving and adding value to research data throughout its lifecycle, according to legal, statutory, ethical and funding body requirements.
- Covers the planning, collecting, organising, managing, storage, security, backing up, preserving, and sharing of your data.



Data Life Cycle

WHAT IS RESEARCH DATA?

all data which is created by researchers in the course of their work, and for which the institution has a curatorial responsibility for at least as long as the Code and relevant archives/record keeping Acts require

&

third-party data which may have originated within the institution or from elsewhere

ANDS Guide: What is research data?
http://ands.org.au/guides/what-is-research-data.html

TYPES OF RESEARCH DATA

- Will vary by discipline
- Digital or analogue
- Numerical, descriptive, visual, audio
- Raw or analysed
- Experimental or observational

WHY HAS UQ LIBRARY ESTABLISHED A RDM SERVICE?

- Global trend towards "open data" and open access, and expectation that publicly funded research is accessible.
- Australian Code for the Responsible Conduct of Research
- UQ Research Data Management Policy
- ARC and other Funding agencies
- Journals require datasets to be submitted with publications

POLICY

The Code, UQ, Funders, Journals

AUSTRALIAN CODE FOR THE RESPONSIBLE CONDUCT OF RESEARCH
AKA
"THE CODE"

Authored by: Australian Government; National Health and Medical Research Council and Australian Research Council

Section 2: Management of Research Data and Primary Materials

Responsibilities of Researchers:

- 2.5 Retain research data and primary materials
- 2.6 Manage storage of research data and primary materials
- 2.7 Maintain confidentiality of research data and primary materials

https://www.nhmrc.gov.au/guidelines-publications/r39

UQ RESEARCH DATA MANAGEMENT POLICY 4.20.06

- Data must be stored, retained, documented and/or described, made accessible for use and reuse, and/or disposed of, according to legal, statutory, ethical and funding body requirements.
- Ownership
- Storage
- Retention
- Disposal
- Access

http://ppl.app.uq.edu.au/content/4.20.06-research-data-management

FUNDING AGENCY REQUIREMENTS:
AUSTRALIAN RESEARCH COUNCIL (ARC)

Researchers […] have an obligation to care for and maintain research data... and **strongly encourages** the depositing of data arising from a Project in an appropriate publicly accessible subject and/or institutional repository

ARC Funding Rules:
http://www.arc.gov.au/funding-rules

Grant applications must outline how they plan to **manage research data.**

Final reports must outline how the research data was made **publicly accessible.**

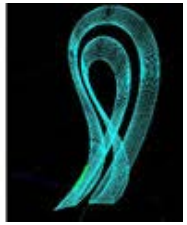
FUNDING AGENCIES WITH RDM REQUIREMENTS

NSF

Australian Government
National Health and Medical Research Council
NHMRC

NATIONAL INSTITUTES OF HEALTH

EPSRC
Engineering and Physical Sciences Research Council



BILL & MELINDA GATES foundation

MRC Medical Research Council

wellcome trust

PUBLISHERS & JOURNALS WITH DATA AVAILABILITY POLICIES

PLoS
PUBLIC LIBRARY of SCIENCE

nature

PNAS

blood

THE ROYAL SOCIETY

The Econometrics Journal

FAIR DATA PRINCIPLES

- In a data intensive research environment it is important to facilitate knowledge discovery by making data discoverable, accessible, analysable by humans and machines.
    - Findable
    - Accessible
    - Interoperable
    - Reusable

FORCE11
The Future of Research Communications and e-Scholarship

SO…

- Your research data is:
  - Important
  - Expected to be shared
  - Expected to be made available
  - That's why you need to make sure your research data is kept in good shape – well managed

WHAT ABOUT THE RISKS?

CONSIDER

- Physical damage – fire, flood, etc
- Corrupt files
- Lost or stolen devices, USB's, laptops
- Sensitive data falling into the wrong hands – commercial in confidence, defence, media, personal information, health, culturally sensitive

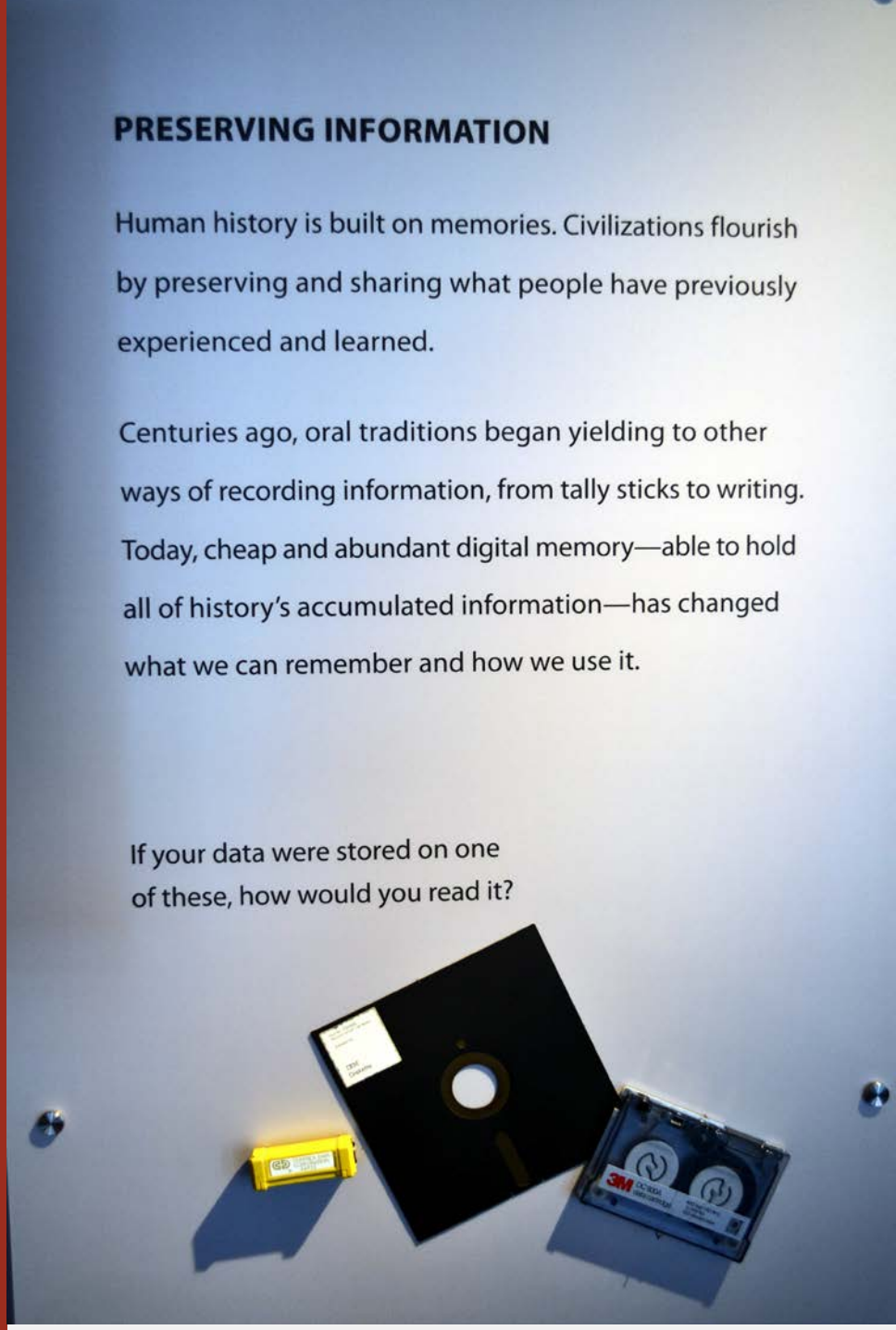
What plans do you have in place to avoid this?

DATA MANAGEMENT PLAN (DMP)

DMP, Best Practice

WHY MANAGE RESEARCH DATA?

- In case you need the data later
- To share your data
- To get credit for producing it
- To maintain the integrity of your data
- To meet obligations to funders or other institutions/organisations



DMP?

A document that describes how you will

collect, organise, store, secure, preserve, back-up, access & reuse

your data

ESSENTIAL ELEMENTS IN A DMP

- Metadata
- Ownership
- Storage
- Retention
- Disposal
- Access

To match UQ Research Data Management Policy requirements

Take a moment to think about your research

- **Talk with your neighbour** and discuss the types of data you generate, where your data is stored and how it is organized.
- If you were asked to share your data with another researcher would they be able to make sense of your data?
- If you needed to locate your data files from 5 years ago, how easy would they be to find and use?

METADATA

WHAT IS METADATA?

"Metadata is structured information that describes, explains, locates, or otherwise makes it **easier to retrieve, use, or manage an information resource**.

Metadata is often called data about data or information about information"

NISO, Understanding Metadata 2004; 1

WHAT IS METADATA?

- Metadata is important for:
  - Discovery (title, keywords, project description)
  - Evaluation (methods, dates, etc.)
  - Reuse (information on software/hardware required, access conditions, etc.)

IMAGINE YOU ARE DESCRIBING YOUR WORK FOR A RESEARCHER 20 YEARS FROM NOW

- **What** it is and what is it for
- **Why** it's different or important
- **How** it relates to existing data
- The **aims** of the research project

METADATA TYPES

Descriptive

About the project
About the data

Structural

Describes how items relate to each other

Administrative

To manage the dataset

ESSENTIAL ELEMENTS IN METADATA

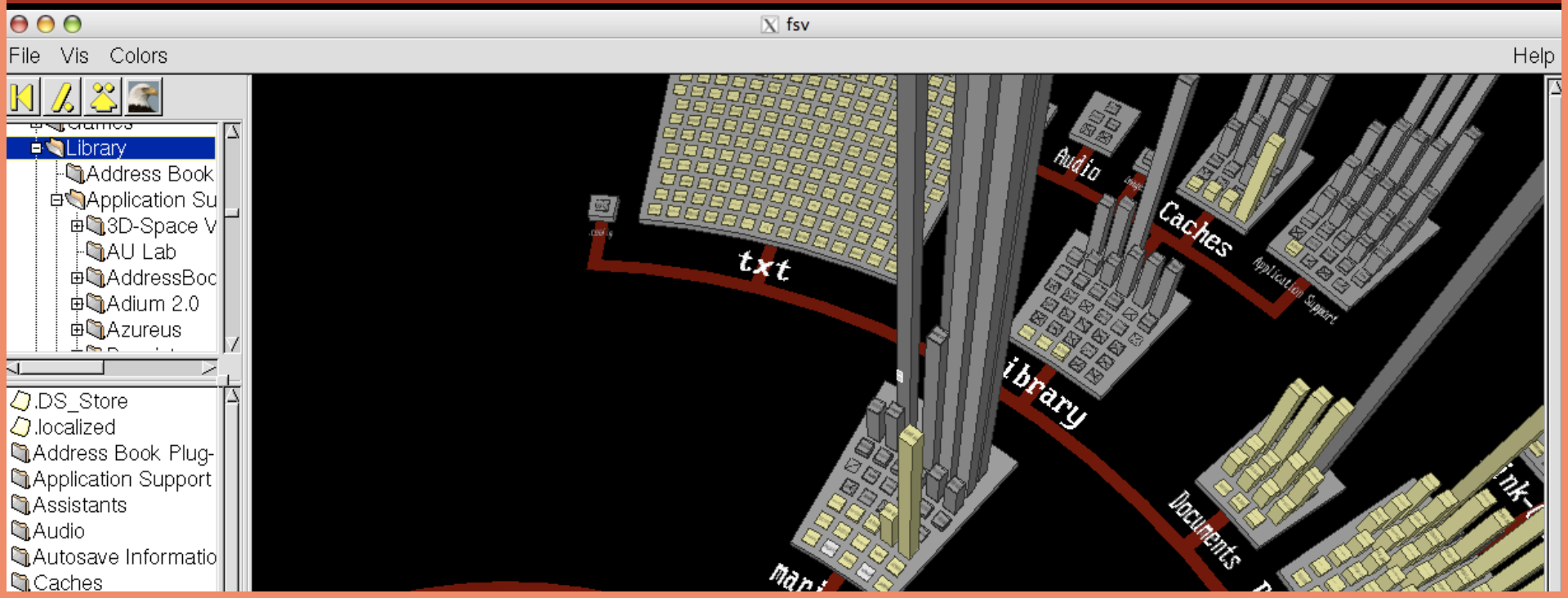
Metadata must be gathered ***as the research is conducted*** (and not added in as an afterthought).

Wellcome Trust
https://goo.gl/zRwSmT

FILES & FORMATS

MANAGING FILES & FORMATS

- Improves efficiency
- Minimises risk of loss or corruption
- Improves data sharing and collaboration
- Ease of re-use



FILES & FORMATS: THINGS TO CONSIDER

- File organisation structure
- File naming conventions
- File formats
- Version control

FILE NAMING CONVENTIONS

- Establish what information to include in file names & **be consistent**
- Content:
  - Date – follow YYYY-MM-DD (ISO 8601)
  - Site
  - Purpose
  - Analysis
  - Researcher
  - Version
- Format:
  - Using underscore or hyphen
  - <25 characters?

Example:

YYYMMDD_site_sampleNum

20140422_PikeLake_03

FILE FORMATS

Best practice requires that data be stored in formats that are most suitable for long term retention and accessibility.

- Unencrypted
- Uncompressed
- Non-proprietary/patent-encumbered
- Open, documented standard
- Standard representation (ASCII, Unicode)

FILE FORMATS

Type	Recommended	Avoid for data sharing
Tabular data	CSV, TSV, SPSS portable	Excel
Text	Plain text, HTML, RTF PDF/A only if layout matters	Word
Media	Container: MP4, Ogg Codec: Theora, Dirac, FLAC	Quicktime H264
Images	TIFF, JPEG2000, PNG	GIF, JPG
Structured data	XML, RDF	RDBMS

Further examples: http://www.data-archive.ac.uk/create-manage/format/formats-table

VERSION CONTROL

UQ pulls study on lack of evidence

THE AUSTRALIAN | SEPTEMBER 04, 2013 12:00AM

SAVE

Sarah Elks
Queensland political reporter
Brisbane

Follow @sarahelks

THE University of Queensland has been forced to ask a respected academic journal to retract a 2011 study on Parkinson's disease because there is no evidence that the research was ever conducted.

The Australian understands a whistleblower lodged a complaint about the study, whose lead author was former UQ professor Bruce Murdoch, an internationally renowned expert on speech and language disorders.

UQ vice-chancellor Peter Hoj said the university had asked the European Journal of Neurology to retract the paper, and it had agreed, on the grounds "no primary data can be located, and no evidence has been found that the study described in the article was conducted".

VERSION CONTROL

Decide:

- how many versions,
- which versions to keep,
- for how long
- how to organise versions

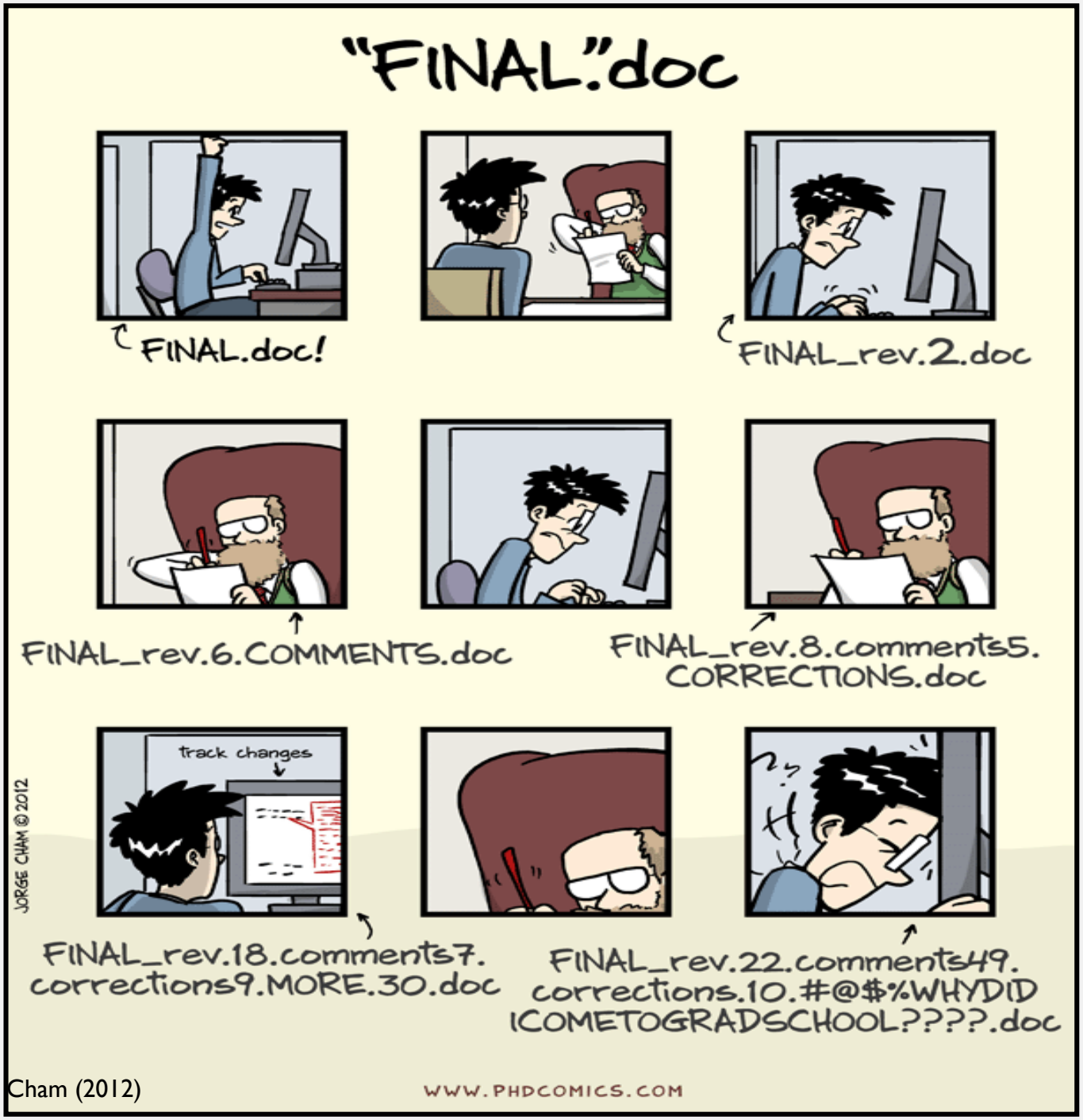
Record:

- Status – draft, interim, final, internal external
- Changes and corrections
- Relationships between files and versions

Document:

- before any data is collected or generated

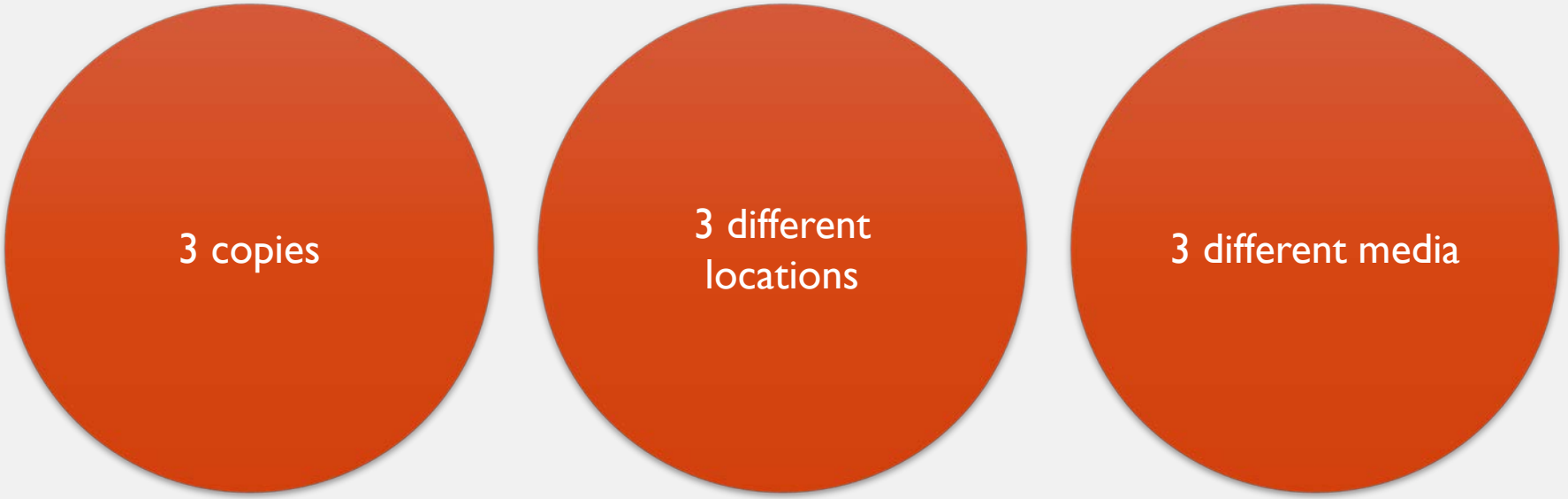
Maintain a SINGLE master copy



Final.doc by Jorge Cham (2012)

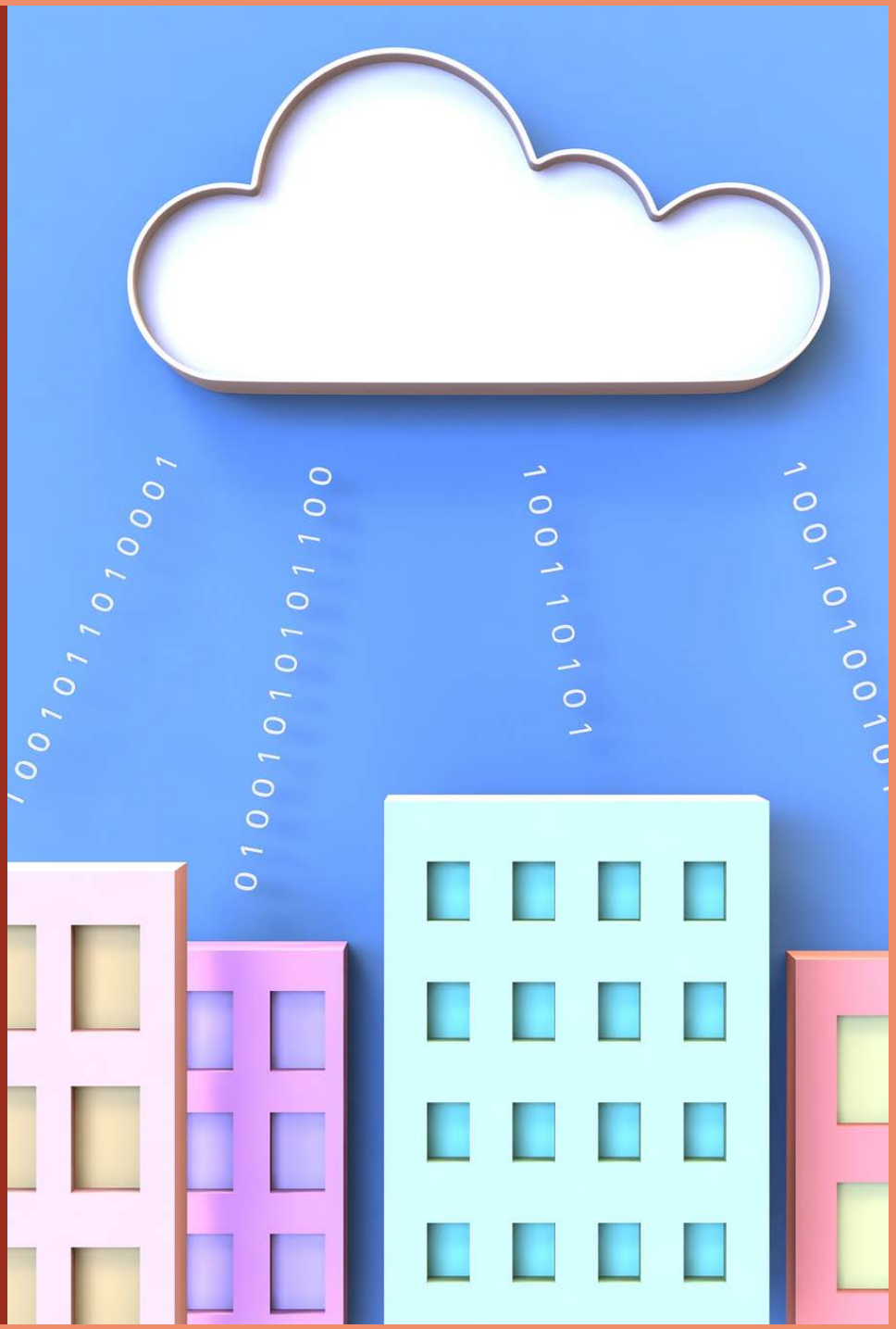
DATA STORAGE

ACTIVE DATA



CONSIDER

- Data type
  - Active data
  - Archive data
- Options
  - Internal
  - External
  - Desktop
  - Cloud



STORAGE: WHAT TO CONSIDER

- How much data you have?
- How often does it need to be accessed and by who (external or off-site collaborators)?
- Is it confidential or sensitive?
- Is it secure?
- Is it backed-up?

STORAGE OPTIONS (ACTIVE)

- Your own drive (PC, server, flash drive, etc.)
- Departmental drive
- University file stores
- Somebody else's drive (collaborating institution)
- CloudStor+ (AARnet) – up to 100 GB
- QRISCloud (QCIF) – large datasets >100GB

STORAGE OPTIONS (ARCHIVED)

UQ ESPACE

- Publishes metadata record for your dataset
- Storage of datasets
- DOI's minted
- Open and mediated access to datasets
- Harvested by Research Data Australia
- Links your publications to your data

OTHER DATA REPOSITORIES

- Dryad – ecology and evolutionary biology, basic and applied biosciences
- Pangaea – earth sciences, georeferenced data
- Gene Expression Omnibus
- Crystallography Open Database
- TERN
- Figshare
- Australian Data Archive
- Registry of Research Data Repositories www.re3data.org

RETENTION

WHAT TO KEEP?

Select what data to keep based on:

- What *has* to be kept (e.g. data underlying publications)?
- What can't be recreated (e.g. environmental recordings)?
- What is potentially useful to others?
- What has scientific, historic or cultural value?
- What must be destroyed?

HOW LONG SHOULD I KEEP IT?

- Data must be kept for statutory periods
- 'Challenged' research data must be kept
- Data for destruction must be destroyed permanently

Queensland State Archives University Sector Retention and Disposal Schedule

http://www.archives.qld.gov.au/Recordkeeping/RetentionDisposal/Schedules/Pages/UniversitySectorQDAN601.aspx

ACCESS & SHARING

WHY SHARE YOUR DATA?

Encourages validation of research methods & results

Provides opportunities for new collaborations

Increases the visibility and potential impact of research

Reduces cost of duplicate data collection…data can be reused in different disciplines

TO CONSIDER…

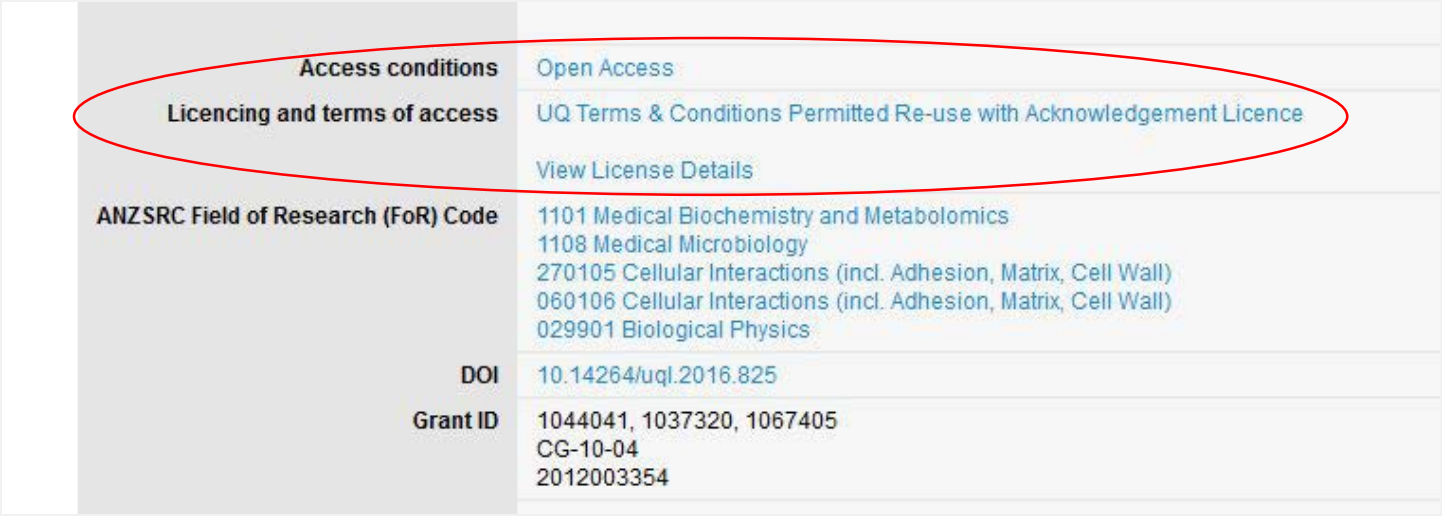
- What obligations or requirements from funders, publishers, your university?
- What's the purpose of access or sharing: to validate the data or to re-use the data for a new project.
- Where? Consider what subject repositories, data centres and structured databases are available?
- Who is responsible for releasing data? Head of School? Establish a Data Access Committee? Individual researcher – but what if they leave UQ?
- Do you have consent for sharing?
- Do any licences you've signed permit sharing?
- Is your data in suitable formats?
- Ethics? – design your research so that it can be shared later.
- Submit your data for publication at time of manuscript submission

LICENSE YOUR DATA

- One of the essential ingredients of reusable data is clarity of reuse permissions, terms, and conditions.
- Prospective re-users need to know exactly what they can do with the data. Those conditions and permissions should be explicit.
- For copyright datasets, use an established and open license

UQ LICENSING AND TERMS OF ACCESS

- UQ General Usage Terms and Conditions for data publishing in eSpace



Access conditions	Open Access
Licencing and terms of access	UQ Terms & Conditions Permitted Re-use with Acknowledgement Licence View License Details
ANZSRC Field of Research (FoR) Code	1101 Medical Biochemistry and Metabolomics 1108 Medical Microbiology 270105 Cellular Interactions (incl. Adhesion, Matrix, Cell Wall) 060106 Cellular Interactions (incl. Adhesion, Matrix, Cell Wall) 029901 Biological Physics
DOI	10.14264/uql.2016.825
Grant ID	1044041, 1037320, 1067405 CG-10-04 2012003354

CREATIVE COMMONS LICENCES

Attribution-NonCommercial-ShareAlike
CC BY-NC-SA

Attribution-ShareAlike
CC BY-SA

Attribution-NonCommercial
CC BY-NC

Attribution-NoDerivs
CC BY-ND

Attribution-NonCommercial-NoDerivs
CC BY-NC-ND

Attribution
CC BY

http://creativecommons.org/choose

EXAMPLE OF CC LICENCES APPLIED TO DATA SETS

All these data are discoverable via the TERN Data Discovery Portal and also accessible directly from facility repositories

Data collection name Description	Licensing*	Acknowledgements
Howard Springs OzFlux tower site Data from the Howard Springs OzFlux flux station in the Black Jungle Conservation Reserve, Northern Territory. Temporal coverage: From 01-01-2001 to 13-05-2013 *Flux data from an additional 22 OzFlux sites are also available for download via the TDDP*		Jason Beringer, University of Western Australia
Geraldton Beach (Video Obs) - Beach Observation 2 min video clip data from a camera at Geraldton surf lifesaving club on Tarcoola beach. Temporal coverage: From 04-01-2010 to 14-09-2011 *Data from 19 other sites around Australia are also available for download via the TDDP*		CoastalCOMS Pty Ltd
Victorian Alpine Plot Network (Phenology Studies): Vegetation (Insect flower visitors) Data, South-east Highlands, Australia, 2013-2014 Phenology data sampled on the same dates, three times a year at transects situated at an altitude of 1400 m to 1880 m in LTERN's Victorian Alpine Plot Network. Temporal coverage: From 2013 to 2014		Dr Ary Hoffmann, The University of Melbourne

*Need help with licensing?

SENSITIVE DATA

Sensitive data are data that can be used to identify an individual, species, object, process, or location that introduces a risk of discrimination, harm, or unwanted attention. Under law and the research ethics governance of most institutions, sensitive data cannot typically be shared in this form, with few exceptions.

ANDS Publishing and sharing sensitive data guidebook
http://www.ands.org.au/guides/sensitivedata

SENSITIVE DATA

Human medical/health research

Research with children and young adults

Research with people with learning difficulties

Research within organisations or the workplace

Research into crime

Internet research

Ecological data

SHARING SENSITIVE DATA?

A person's identity can be disclosed from:

- Direct identifiers, e.g. name, address, postcode, telephone number
- Indirect identifiers that, when linked with other publicly available information sources, could identify someone, e.g. workplace, occupation, salary or age
- Special attention maybe needed for:
  - Relational data, where relations between variables in related datasets can disclose identifies
  - Geo-referenced data, where identifying spatial references such as point co-ordinates also have a geographical value

SHARING SENSITIVE DATA?

- Anonymise digital data by removing direct and indirect identifiers, or aggregating data
- Separating disclosable from non-disclosable data by obscuring, removing or hiding individual fields, records or data files
- Remove geo-referenced data or point coordinates
- Encrypting data – software that can help you:
  - Open Source GnuPG, or
  - Commercial Pretty Good Privacy (PGP)

ANDS Publishing and sharing sensitive data guidebook: http://www.ands.org.au/guides/sensitivedata

FINDING PUBLISHED DATA

- Use the Library's Research Data Guide to find sources of publicly accessible data
- http://guides.library.uq.edu.au/how-to-find/research-data
- Research Data Australia
- Other data repositories

Finding open, accessible research data sets can improve efficiency, speed up the research process and reduce costs and duplication of effort

DATA SHARING AND MANAGEMENT SNAFU IN 3 SHORT ACTS

DATA CITATION

Get recognised for your research

WHY CITE DATA?

- Receive (and give) acknowledgement and recognition for datasets
- Increases visibility of research
- Establish new collaborations
- UQ eSpace adds DOI's to your data records

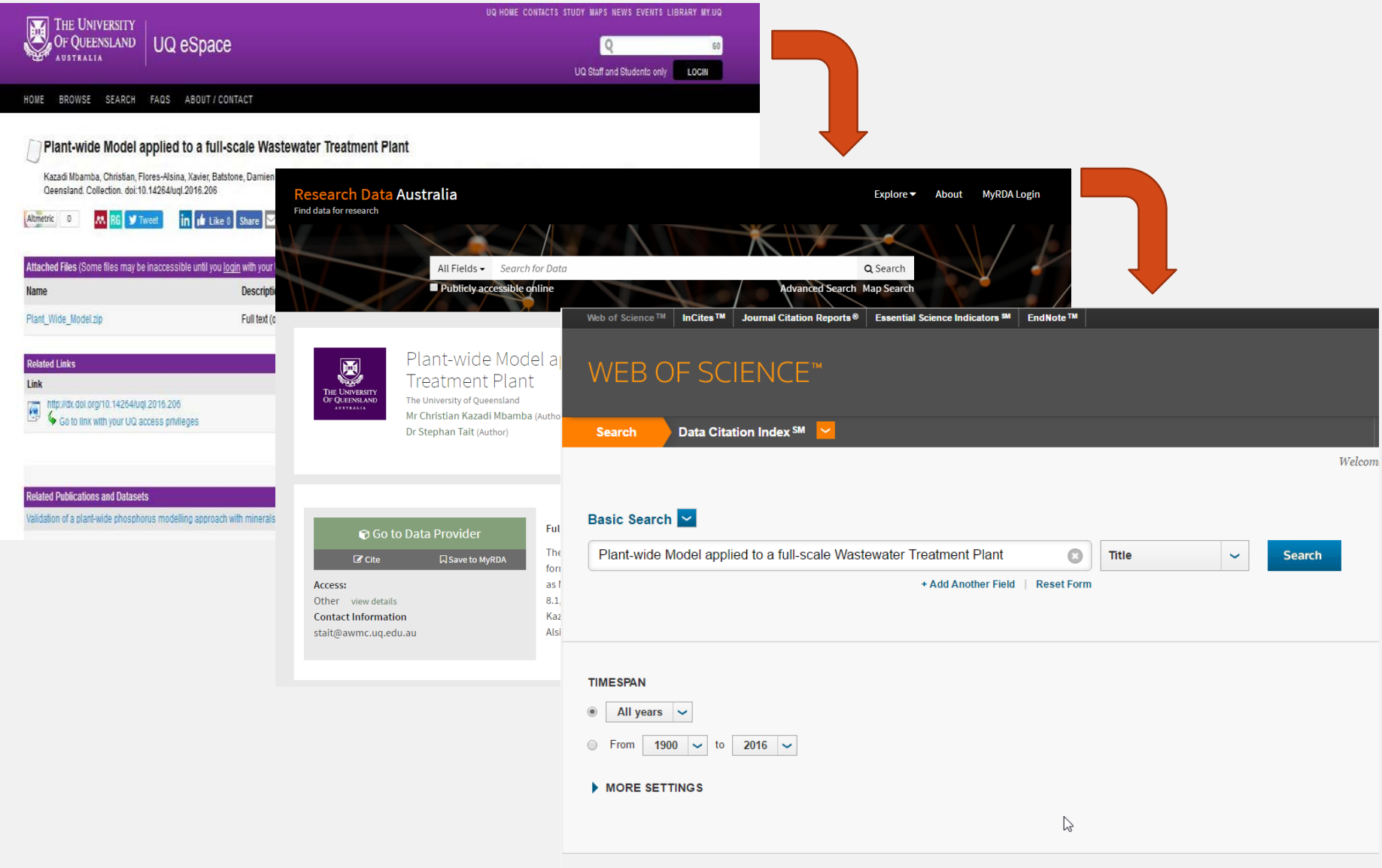
HOW TO CITE A DATASET

- DOI Citation Formatter (via CrossCite.org)
- ANDS Guidelines on Data Citation

Hanigan, Ivan (2012): Monthly drought data for Australia 1890-2008 using the Hutchinson Drought Index. The Australian National University Australian Data Archive. http://doi.org/10.4225/13/50BBFD7E6727A

Pew Hispanic Center. (2008). *2007 Hispanic Healthcare Survey* [Data file and code book]. Retrieved from http://www.pewhispanic.org/2007/09/23/2007-hispanic-healthcare-survey/

UQ ESPACE (& OTHERS) → RESEARCH DATA AUSTRALIA → TR DATA CITATION INDEX



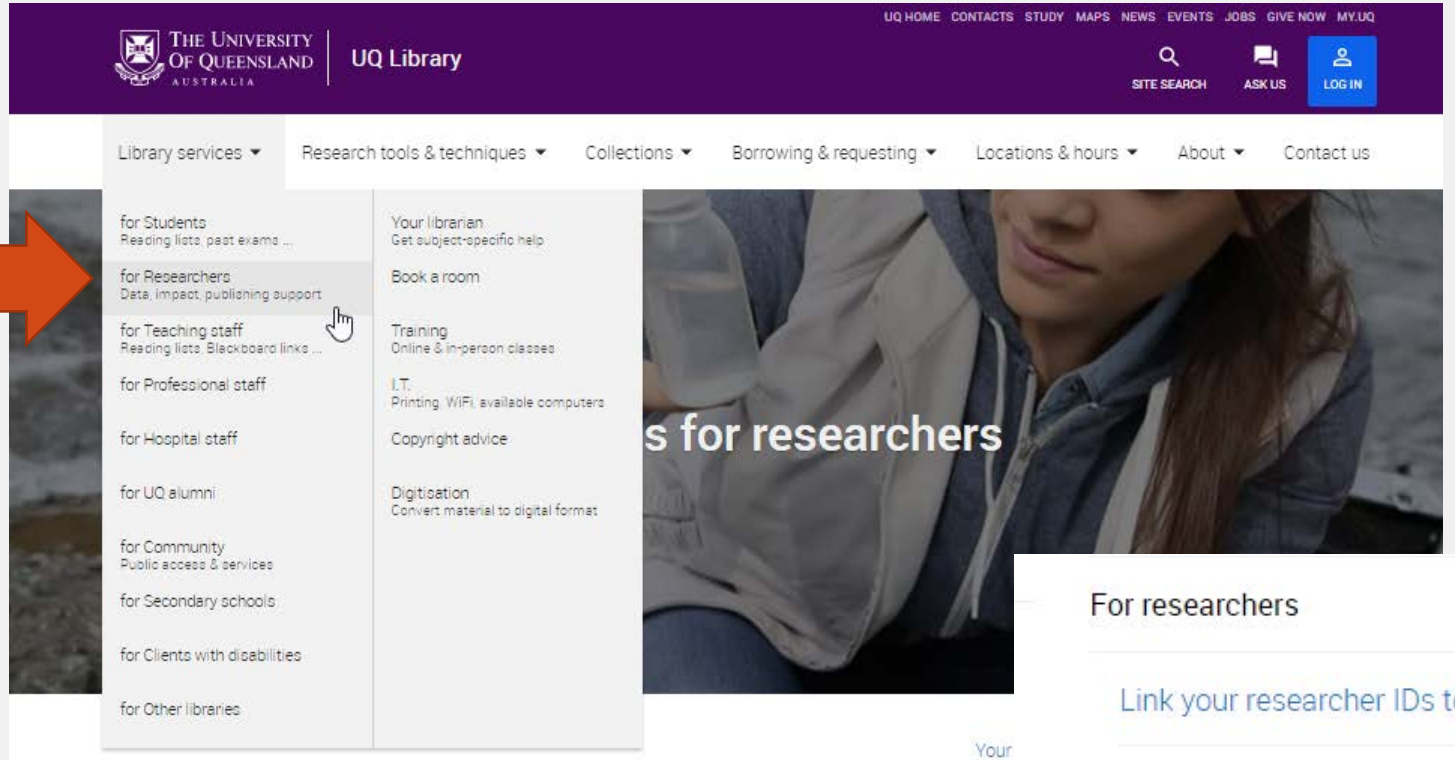
TOOLS, RESOURCES

UQ LIBRARY RESOURCES

- Research Data Management Guide
  - http://guides.library.uq.edu.au/how-to-find/research-data

- Online data management planning tool
  - https://dmponline.app.uq.edu.au/

  - Access both via Library Home page/Library Services/ For Researchers



How can we help you do your best research?

1. I'd like to contact my librarian
2. I need help finding resources
3. Please order something for me
4. I'd like some training
5. I've got research data to manage
6. I need advice on publishing
7. Is my research making an impact?
8. Can UQ store and promote my research?

1. I'd like to contact my librarian

Your librarian can provide personal help with:

- Research consultations
- Literature searches and reviews
- Database training - refresher and advanced
- Library tours and orientation

For researchers

- Link your researcher IDs to UQ eSpace
- Researcher identifiers and profiles guide
- Getting published guide
- Open access guide
- Research data management guide
- Research impact and metrics guide
- UQ Author Statistics
- Request an ISBN or ISSN

Research data management: Get started

Search this Guide Go!

Research data management covers the planning, collecting, organising, managing, storage, security, backing up, preserving, and sharing your data.

Welcome

This guide provides information about managing research data. It is intended for staff and students at the University of Queensland.

What is research data management?

Effective research data management is a cornerstone of the responsible conduct of research.

It ensures that research data are managed according to legal, statutory, ethical and funding body requirements. Research Data Management covers the planning, collecting, organising, managing, storage, security, backing up, preserving, and sharing your data. Research data represents significant value to researchers and the University of Queensland, and good stewardship of research data is necessary to validate the outcomes of research, and maintain the integrity of research results.

Online Data Management Planning Tool

- Online Data Management Planning Tool

Use this tool to write, save, share, and manage your Data Management Plans. User Guide

EXAMPLE PLANS

- Sample DMPs on UK Digital Curation Centre site
  http://www.dcc.ac.uk/resources/data-management-plans/guidance-examples

- Sample DMPs on DMPTool site
  https://dmptool.org/public_dmps

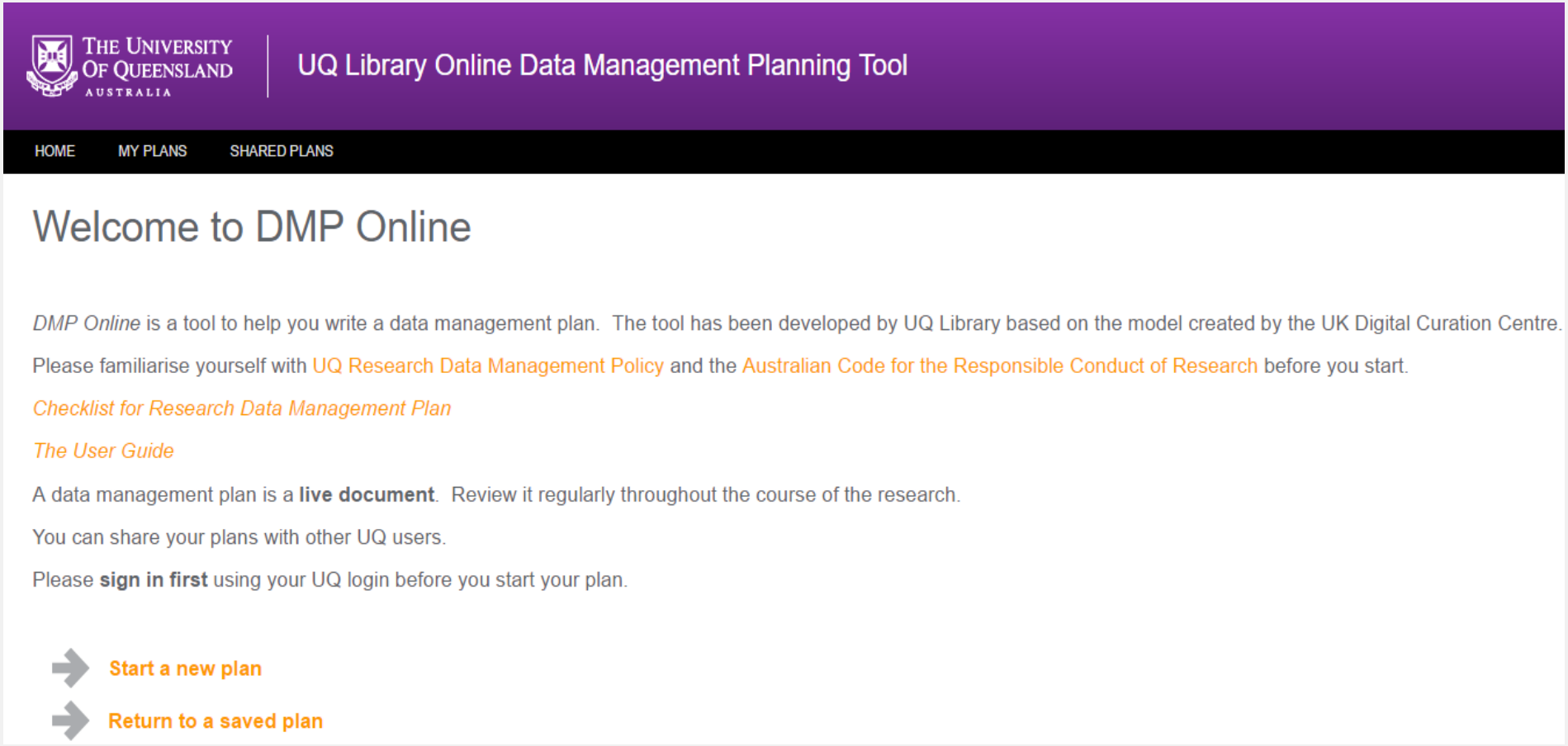
- Rural Economy & Land Use (RELU) programme examples
  http://relu.data-archive.ac.uk/data-sharing/planning/examples

SAMPLE RDM PLAN FOR ARC GRANT

- Data collected will be managed in line with the Australian Code for the Responsible Conduct of Research.
- The Data Management Plan for the proposed project will include Project Description, Context, and Data Capture Formats. We will use the services offered by the UQ Library's e-Space relating to capturing metadata about data, organising and storing data, long-term preservation, ethics and intellectual property and arrangements around data sharing and re-use.
- In addition, all data will be made available publicly through e-Space in order to encourage replication studies or additional analyses of the data collected by us.

HANDS ON

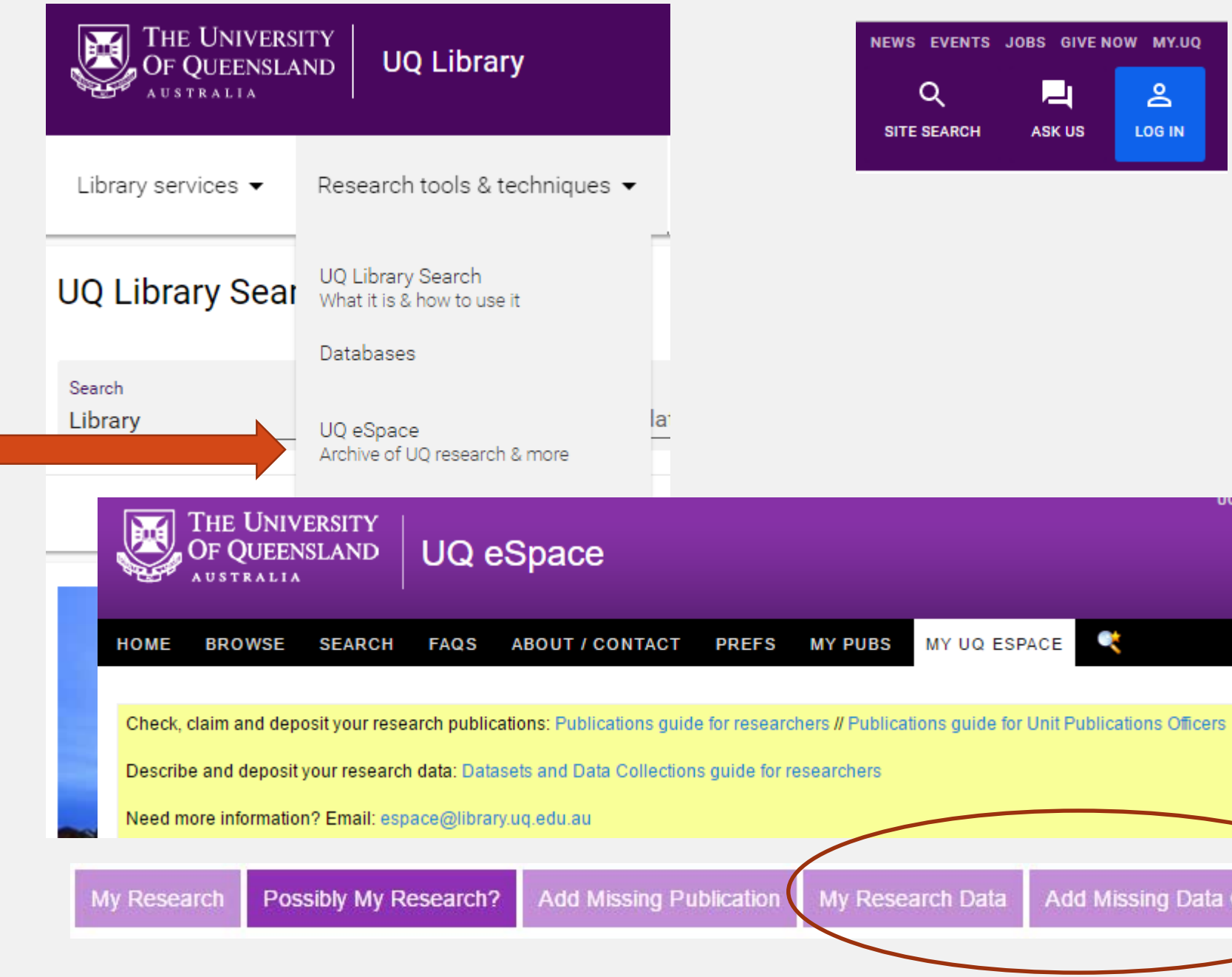
DMP ONLINE


THE UNIVERSITY OF QUEENSLAND
AUSTRALIA
UQ Library Online Data Management Planning Tool
HOME
MY PLANS
SHARED PLANS
Welcome to DMP Online
DMP Online is a tool to help you write a data management plan. The tool has been developed by UQ Library based on the model created by the UK Digital Curation Centre.
Please familiarise yourself with UQ Research Data Management Policy and the Australian Code for the Responsible Conduct of Research before you start.
Checklist for Research Data Management Plan
The User Guide
A data management plan is a live document. Review it regularly throughout the course of the research.
You can share your plans with other UQ users.
Please sign in first using your UQ login before you start your plan.
Start a new plan
Return to a saved plan

PUBLISH YOUR DATA UQ E-SPACE

- UQ eSpace makes your research data visible via
  - search engines such as **Google**,
  - **Research Data Australia**.
  - Thomson Reuters' **Data Citation Index** database.

- **Metadata records** describe a researcher's data under their My UQ eSpace profile in eSpace ('My Research Data' tab, allowing researchers to build an index of their research data, and count the number of times they have been viewed and downloaded, therefore contributing to their research profile.


THE UNIVERSITY OF QUEENSLAND
AUSTRALIA
UQ Library
NEWS EVENTS JOBS GIVE NOW MY.UQ
SITE SEARCH
ASK US
LOG IN
Library services
Research tools & techniques
UQ Library Search
What it is & how to use it
Databases
UQ eSpace
Archive of UQ research & more
Search
Library
THE UNIVERSITY OF QUEENSLAND
AUSTRALIA
UQ eSpace
HOME BROWSE SEARCH FAQS ABOUT / CONTACT PREFS MY PUBS MY UQ ESPACE
Check, claim and deposit your research publications: Publications guide for researchers // Publications guide for Unit Publications Officers
Describe and deposit your research data: Datasets and Data Collections guide for researchers
Need more information? Email: espace@library.uq.edu.au
My Research
Possibly My Research?
Add Missing Publication
My Research Data
Add Missing Data Collection